

The background of the entire page is an abstract composition. It features large, overlapping, organic shapes in shades of brown, tan, and purple, resembling stylized leaves or flowing liquid. Overlaid on these shapes are numerous small, bright yellow and orange star-like points, some of which are arranged in faint, curved trails, suggesting a cosmic or digital theme. The overall color palette is warm and textured.

Center for Advanced Computing Research

building collaborations
to extend
the frontiers of
science and engineering
through advanced computing research



On the cover: Visualization of the Rayleigh-Taylor fluid flow instability.

Rayleigh-Taylor instabilities occur whenever a fluid of lower density is accelerated into an adjacent fluid of higher density. Such instabilities lead to turbulent mixing at the interface between the fluids. The instability occurs in many applications, such as the collapse of supernovae or the implosion of deuterium capsules by high powered lasers to achieve fusion. Shown in the figure is an isosurface of the density field of the fluid at an intermediate time. The isosurface is intersected with an orthogonal slice immersed in a rendered field of evenly spaced density samples with color representing differing density values. The flow field data used to render this image originates from a large-scale data set generated from high-resolution simulations performed by Dr. Andrew Cook on the ASCI Blue Pacific Teraflops-scale platform at the Lawrence Livermore National Laboratory. This work was carried out in collaboration with Professor Paul Dimotakis and Dr. Ron Henderson of the compressible turbulence research group at the Caltech ASCI alliance center. Visualizations such as the one shown here are of great value in providing insight into the mechanisms of turbulent mixing driven by the instability. This visualization was produced by Santiago V. Lombeyda of CACR.

CONTENTS

- 1 CACR Mission
- 2 What is Computational Science and Engineering?
- 2 Computation: The Third Research Method
- 3 Scientific and Engineering Applications
- 4 The Computer as a Laboratory
- 6 Extending the Frontiers
 - 6 *Simulating the Dynamic Response of Materials*
 - 6 *Federating the Digital Sky Surveys*
 - 7 *Gearing up for the Data Thunderstorm in Particle Physics*
 - 8 *Applying Oil Painting Techniques to Multivalued Scientific Data Visualizations*
 - 8 *Nuclear Matter Calculations on a Lattice*
 - 9 *An Active Digital Library for LIGO*
 - 10 *Computer Science Research at CACR*
 - 11 *High-Performance Computing in Chemistry: From Very Simple to Very Complex Systems*
- 12 CACR Computing Facilities

ACKNOWLEDGMENTS

Brochure Design: Tina Mihaly, Santiago V. Lombeyda, Juley Sobson
 Image credits: Santiago V. Lombeyda (cover image), Juley Sobson (collage on pages 4-5), Sarah Emery (facilities photos), Tina Mihaly (campus photo)
 Editing: Tina Mihaly, Jim Pool, Paul Messina, Roy Williams, Juley Sobson
 Contributions: Mark Bartelt, Randy Bramley, Julian Bunn, Sharon Brunett, Bruce Char, Scott Fraser, Ed Givelberg, Steve Koonin, Hans-Michael Muller, David Kremers, Aron Kuppermann, Vince McKoy, Dan Meiron, Paul Messina, Tina Mihaly, James Patton, Ron Perline, Roy Williams, Carl Winstead, Mark Wu

CACR Mission

The Center for Advanced Computing Research (CACR) was established at Caltech to foster advances in computational science and engineering. To achieve this goal, our center conducts multidisciplinary, application-driven research in Computational Science and Engineering (CS&E) and participates in a variety of high-performance computing and communications research and development activities. In addition to its research activities and creating large-scale computing facilities, CACR acts as a catalyst for the advancement of information and computing technologies at Caltech and its Jet Propulsion Laboratory.

CACR aims to enable breakthroughs in computational science and engineering by

- ♦ following an applications-driven approach to computational science and engineering research,
- ♦ conducting multidisciplinary research on leading-edge computing facilities,
- ♦ providing an intellectual environment that cultivates multidisciplinary collaborations, and
- ♦ harnessing new technologies to create innovative large-scale computing environments.

CACR's library in the commons area of Caltech's Powell-Booth Laboratory for Computational Science. In 1998, Caltech began extensive renovations of the Willis H. Booth Computing Center to accommodate computational science for the 21st century. The renovations were funded by generous grants from the Charles Lee Powell Foundation, the National Science Foundation, Caltech, and the Booth Ferris Foundation. This interaction space, which was envisioned by CACR Director Paul Messina to be a "grassy meadow," was designed to facilitate collaborations between CACR staff, Caltech researchers, and visiting scientists. The new Caltech facility—which was dedicated on April 2, 1999 and occupied by CACR staff in May 1999—is now known as the Powell-Booth Laboratory for Computational Science.



*In 1995, the Center for
Advanced Computing
Research was
established to ensure
that the California
Institute of Technology
and its Jet Propulsion
Laboratory will be at
the forefront in
computational science
and engineering.*

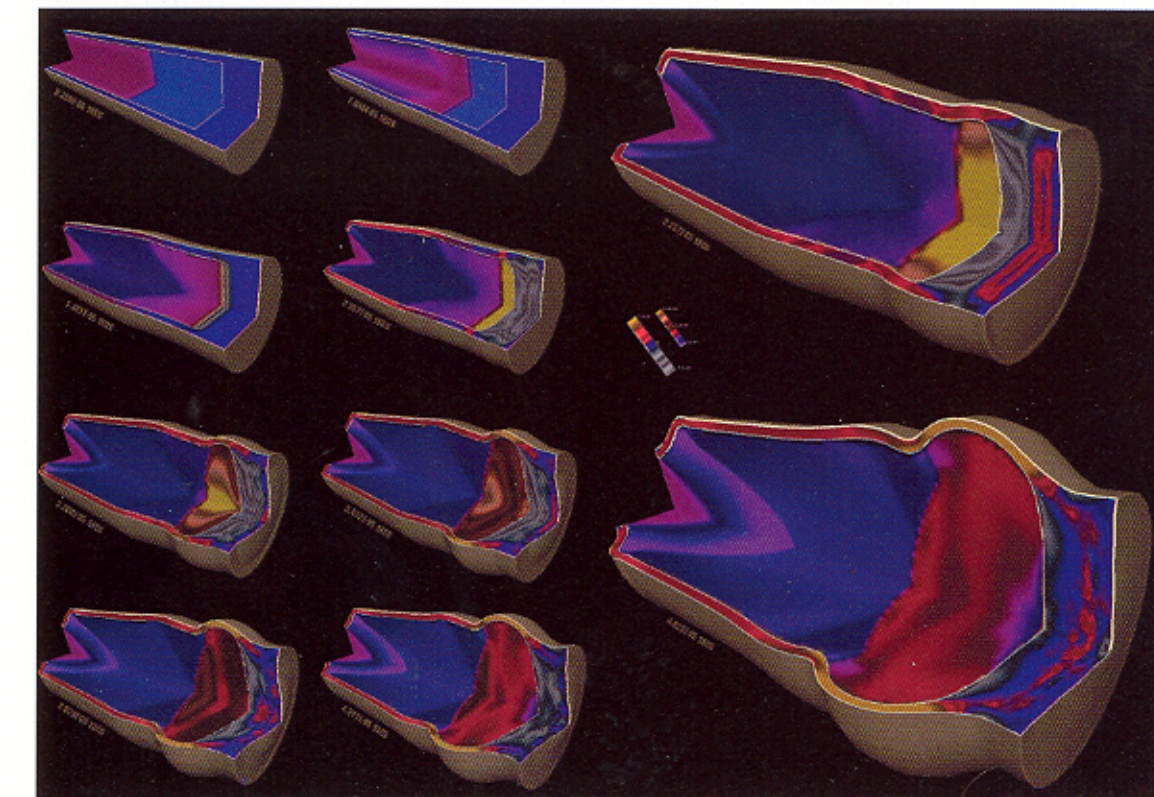
What is Computational Science and Engineering?

Computational Science and Engineering is the practice of computer-based modeling and simulation for the study of scientific phenomena and engineering designs. Computational science and engineering research is multidisciplinary, as it requires:

- ♦ knowledge and methodologies from the application fields, from computer science, and from mathematics;
- ♦ expertise in the use of the technologies that comprise today's computing environments; and
- ♦ the integration of the information technologies into new combinations that provide new capabilities.

Computation: The Third Research Method

Computer-based modeling and simulation are indispensable for gaining a better understanding of many scientific phenomena and engineering designs. In recent years, computation has become the third research methodology, complementing theory and experiment. Today, computing environments and methods for using them have become powerful enough to tackle



Computer-based modeling and simulation are essential tools in a number of fields in which researchers deal with phenomena that cannot be produced easily in a laboratory. An example is the visualization above, depicting a simulation performed in the DOE-supported Caltech virtual test facility (see page 6). In this simulation, a strong shock wave is interacting with a layered metal target. The fluid is simulated with a parallel Eulerian solver, while the solid is simulated with a dynamically adaptive Lagrangian code. The solid-fluid coupling is accomplished via a dynamically updated level set.

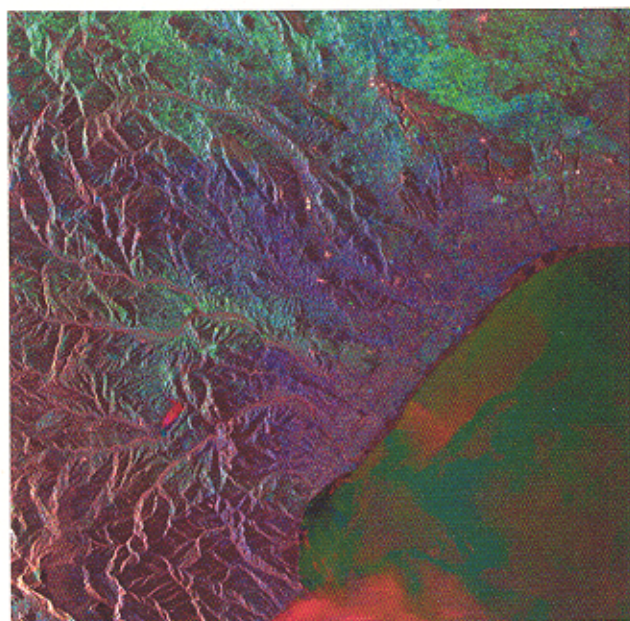
problems of great complexity. An indication of this is that many of the research areas that the California Institute of Technology will focus on during the next decade require state-of-the-art computing capability to accomplish their research agenda.

Computer simulation makes it possible to investigate regimes that are beyond current experimental capabilities and to study

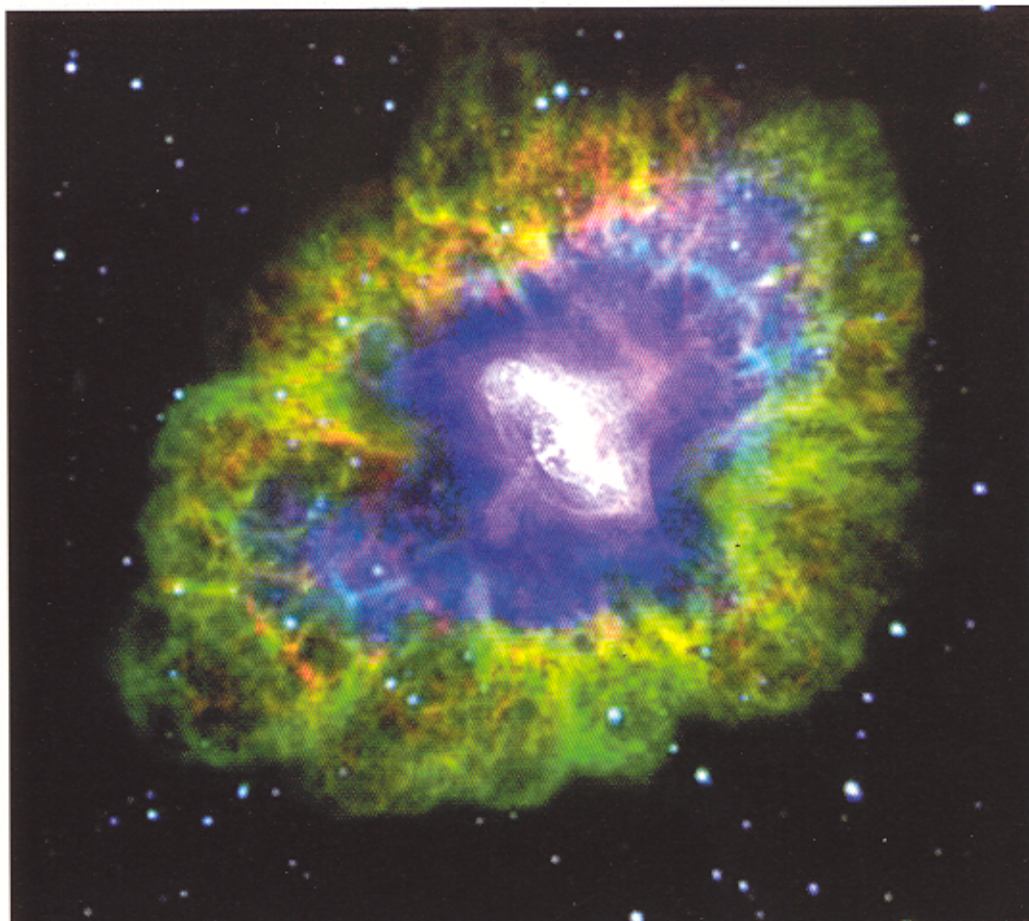
phenomena that cannot be replicated in laboratories, such as the evolution of the universe. In the realm of science, computer simulations are guided by theory as well as experimental results, while the computational results often suggest new experiments and theoretical models. In engineering, many more design options can be explored through computer models than by building physical ones, usually at a small fraction of the cost and elapsed time.

Scientific and Engineering Applications

The large-scale applications that Caltech researchers and collaborators perform on CACR machines span a wide spectrum of scientific and engineering fields,



including astronomy, biology, chemistry, computer science, computational mathematics, earth science, geophysics, fluid dynamics, and physics. Selected computational projects currently in progress at the California Institute of Technology and at collaborating research organizations are summarized on pages 6-10 and in the inserts located in the inside back cover. In addition, a sampling of computational results obtained on state-of-the-art computer systems are featured in the collage on pages 4-5.

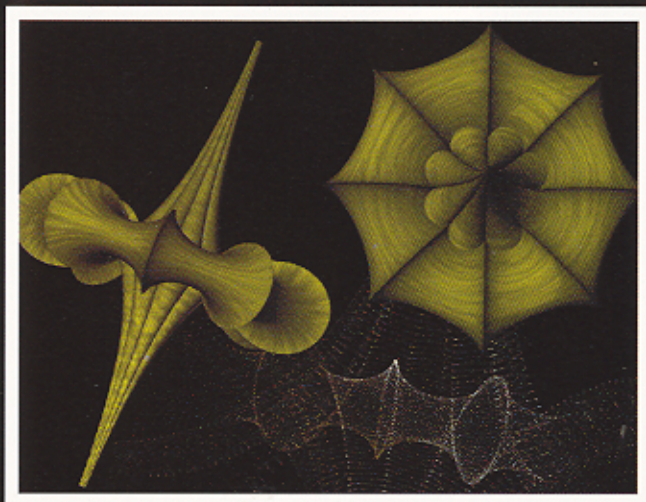


Above: The Crab Nebula in Orion, showing also the central engine. Roy Williams of Caltech CACR constructed this image by overlaying the X-ray image from the Chandra Observatory with an optical image from the Palomar Observatory. The picture illustrates how a new view can be created simply by combining different data sources—the driving concept of the Digital Sky project (see “Federating the Digital Sky” on page 6).

At left: Synthetic Aperture Radar multitemporal image of a part of southern Italy. This image from the Digital Puglia project, produced by Roy Williams of Caltech CACR, was obtained using three monochrome ERS-2 images taken at different times. Red, green, and blue represent the months of May, August, and November, respectively. Digital Puglia is a prototype digital library of remote-sensing data covering the Puglia region of Italy. The system features a web-based interface and fault-tolerant distribution system that can fuse heterogeneous data and metadata. Digital Puglia employs the web-based retrieval mechanism developed in the Synthetic Aperture Radar Atlas project, a joint effort between Caltech and the University of Lecce in Italy (see http://www.cacr.caltech.edu/SDA/digital_puglia.html).

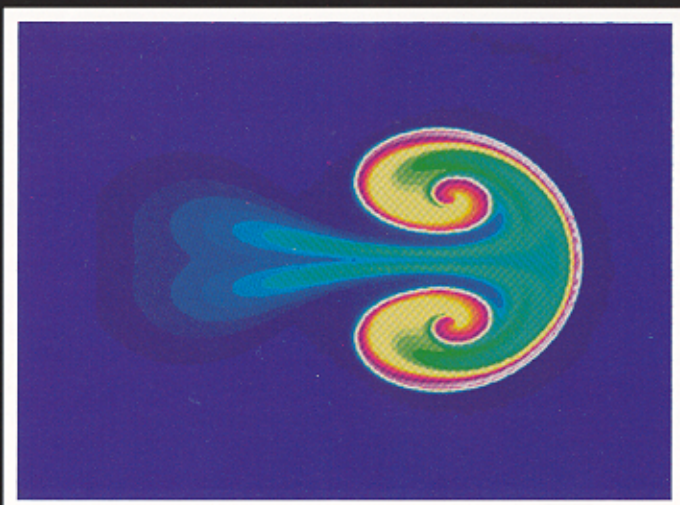
THE COMPUTER AS A LABORATORY

Today, scientists simulate with a computer what they once had to create in a laboratory or on a bench top, or study through theory. Here is a collage of large-scale scientific computations accomplished by using high-performance computer systems and advanced programming techniques. The next generation of computers promises to enable computational problems of even greater size and complexity.

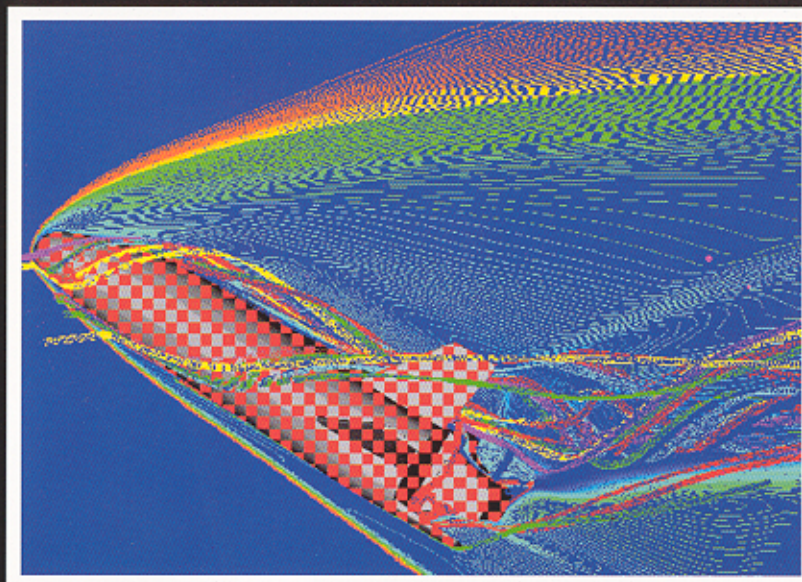


1

(1) This image shows various representations of soliton pseudospherical surfaces, surfaces of special type in geometry. The visualization reveals one surface's periodic structure and self-intersection. The Soliton Explorer Problem Solving Environment developed as part of the Problem Solving Environment project at Drexel has provided tools that reveal structural features previously not well understood in these type of surfaces. (Drexel PSE software group)

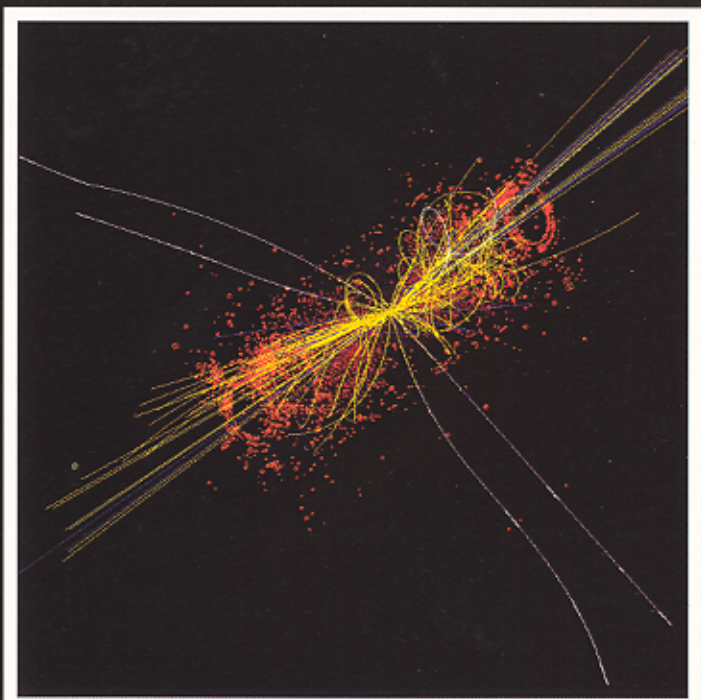


(3) Torodial jet computed by solving incompressible Euler equations. This image reveals the cross-sectional velocity within the jet. (Ron D. Henderson and Dan Meiron, Caltech)



2

(2) This image illustrates the flow over a proposed launch vehicle computed on the CACR HP Exemplar. In this Navier-Stokes solution (angle of attack=35.8 degrees, Mach 11), the lines show the Mach number distribution in the symmetry plane, while ribbons depict some selected stream traces. (Jochem Häuser, Roy Williams, Ralf Winkelmann)

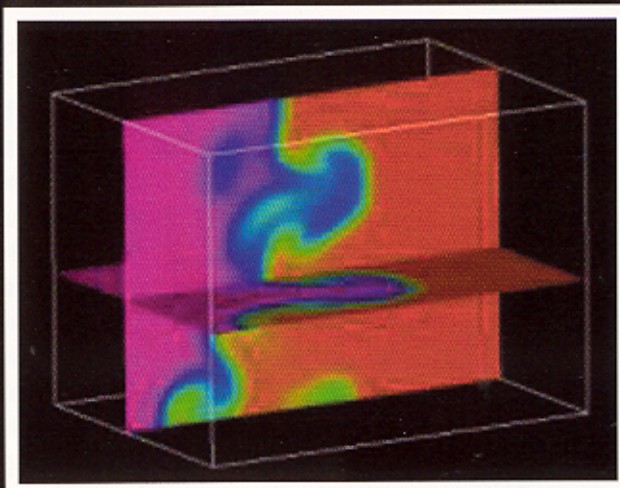


4

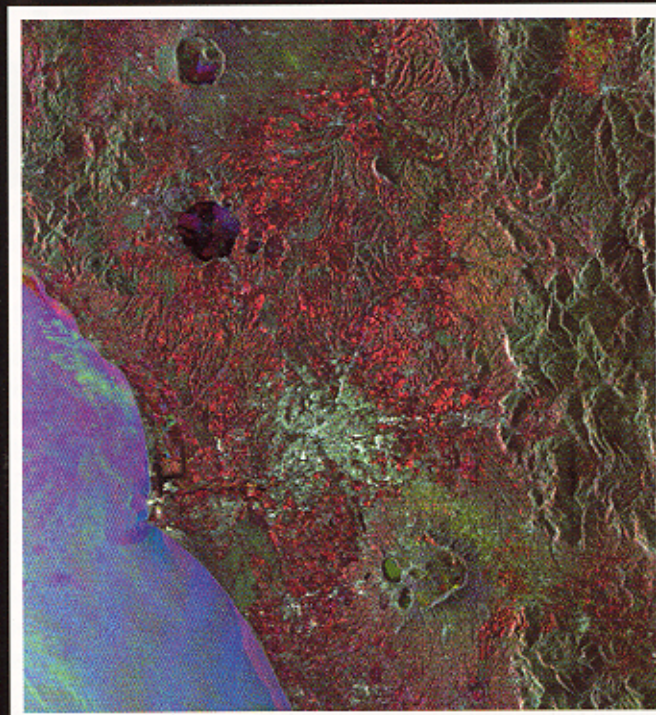
3

(4) A decay of the postulated Higgs boson into four muons, as simulated and captured in the CMS detector. The muon tracks are those appearing in white. Scientists are gearing up for the tremendous volumes of data to be produced by physics experiments of the new century. (Lucas Taylor, Northeastern University)

(5) Richtmyer-Meshkov instability. This simulation was performed by Dr. Ravi Samtaney on the Pittsburgh CRAY T3E. The study of shock-mediated instabilities is a key area of research for the Caltech ASCI Alliance Center (see page 6).

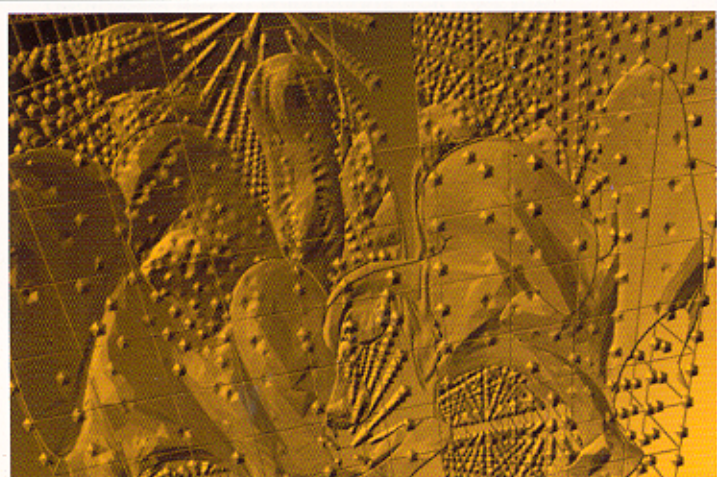


5



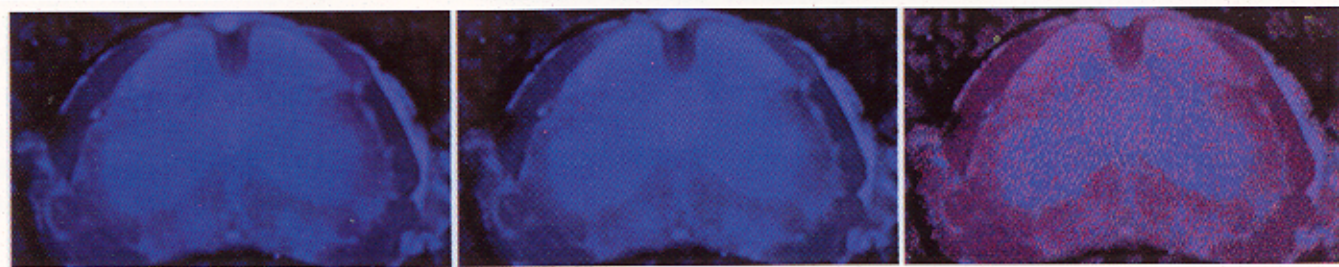
6

(6) Multitemporal SAR image of Rome, Italy, from the Digital Puglia project (see page 3). This image was constructed from three monochrome ERS-1 images taken at different times. (Roy Williams)



7

(7) Visualization of the results of a simulation performed by Dr. Andrew Cook on the ASCI Blue Pacific Teraflops-scale platform at LLNL. (Santiago V. Lombeyda, Caltech CACR)



8

(8) Caltech scientists are applying techniques from oil painting to multivalued scientific data visualizations (see page 8). These images depict the spinal cord of a mouse with experimental allergic encephalomyelitis. From left to right: an underpainting layer, a checkerboard layer composed on the underpainting, and a stroke layer over that. (David H. Laidlaw, David Kremers, Eric T. Ahrens, and Matthew J. Avalos, Caltech)

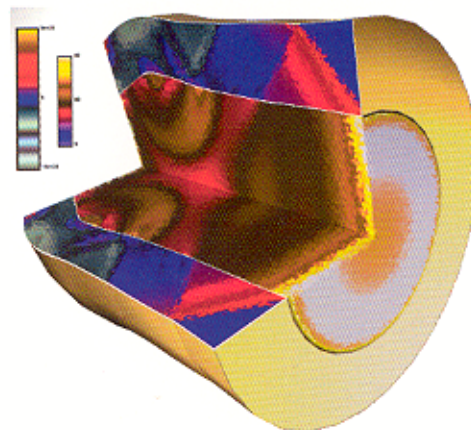
Extending the Frontiers

Simulating the Dynamic Response of Materials

At Caltech, researchers at the Center for Simulating the Dynamic Response of Materials are constructing a virtual shock physics facility in which the full three-dimensional response of a variety of target materials can be computed for a wide range of compressive, tensional, and shear loadings, including those loadings produced by detonation of energetic materials. The effort, which is part of the Academic Strategic Alliances Program of the DOE Accelerated Strategic Computing Initiative, has three objectives:

- ♦ to facilitate computation of a variety of experiments in which strong shock and detonation waves are made to impinge on targets consisting of various combinations of materials,
- ♦ to compute the subsequent dynamic response of the target materials, and
- ♦ to validate these computations against experimental data.

The funded research is centered on the three primary stages required to conduct a virtual experiment in this facility: detonation of high explosives, interaction of shock waves with materials, and shock-induced compressible turbulence and mixing. The modeling requirements are addressed through five integrated research initiatives that guide the key disciplinary activities:



A visualization of the interaction of a detonation wave with a solid enclosing metal cylinder. Such simulations are used to assess the behavior of high explosives. A notable aspect of this simulation is that it integrates two different solvers: a parallel Eulerian fluid mechanics-solver for the high explosive and a Lagrangian solid mechanics solver for the enclosing casing. The simulation was performed using the Virtual Test Facility, the key application being developed by the Center for Simulating the Dynamic Response of Materials at Caltech.

- ♦ modeling and simulation of fundamental processes in detonation,
- ♦ modeling dynamic response of solids,
- ♦ first-principles computation of materials properties,
- ♦ compressible turbulence and mixing, and
- ♦ computational and computer science infrastructure.

Federating the Digital Sky Surveys

Large-area digital sky surveys are a recent and exciting development in astronomical research. The recent large-area surveys in optical, infrared, and radio wavelengths, in combination with the terabytes/teraflops computational resources of the National Partnership for Advanced Computational Infrastructure (NPACI), are providing unprecedented new capabilities for astronomical research. With these capabilities, astronomers will have access to a multi-wavelength "Digital Sky" covering a significant fraction of the real sky.

Participants in the NPACI Digital Sky Project, led by Caltech professor of physics Thomas Prince, are federating the data in the available large-area sky surveys to create a virtual observatory. The "Digital Sky" offers a new perspective of the universe which is statistical and data-focused, in contrast to traditional work with individual stellar objects. In addition, the Digital Sky—and the tools for its exploration—will revolutionize multi-wavelength astronomical studies, both by the sheer increase of the data available, and by providing faster and more sophisticated methods for its analysis.

Among the surveys being undertaken in the Digital Sky project are the Digital Palomar Observatory Sky Survey (DPOSS), the 2-Micron All Sky Survey (2MASS), and the NRAO VLA Sky Survey (NVSS). These surveys alone are expected to yield over one billion sources, and the image data will comprise several tens of

terabytes. By placing these surveys on-line, astronomers will be able to launch sophisticated queries to the catalogs describing the optical, radio, and infrared sources (see example on page 3), and then undertake detailed analysis of the images for morphological and statistical studies of both discrete sources and extended structures. The computing facilities, both at CACR and SDSC, will provide significant capability for pattern recognition, automatic searching and cataloging, and computer assisted data analysis.

Not all the data are literally on-line. The actual databases are so voluminous that they are being stored in HPSS-managed archival storage systems at Caltech and the San Diego Supercomputer Center. However, the Caltech researchers are interfacing the database system to HPSS to enable users of the Digital Sky to issue database queries even for data in the HPSS archive.

After the initial work involving the Palomar, 2MASS, and NVSS surveys, the scientists plan to incorporate other catalogs and surveys in other wavelengths. The Caltech research team is also coordinating with another major survey, the Sloan Digital Sky Survey, and is working toward the creation of standards and conventions that will allow common access of all the surveys in the future. As other surveys come on-line, they will be able to take immediate advantage of the architecture and software developed in the Digital Sky and Sloan projects.

Gearing Up for the Data Thunderstorm in Particle Physics

A data thunderstorm is gathering on the horizon with the next generation of particle physics experiments. The amount of data is overwhelming. Even though the prime data from the CERN Compact Muon Solenoid (CMS) detector will be reduced by a factor of more than 10^7 , it will still amount to over a Petabyte (10^{15} bytes) of data per year accumulated for scientific analysis. The task of finding rare events resulting from the decays of massive new particles in a dominating background is even more formidable. Particle physicists have been at the vanguard of data-handling technology, beginning in the 1940's with eye scanning of bubble-chamber photographs and emulsions, through decades of electronic data acquisition systems employing realtime pattern recognition, filtering and formatting, and continuing on to the Petabyte archives generated by modern experiments. In the future, CMS and other experiments now being built to run at CERN's Large Hadron Collider (LHC) expect to accumulate on the order of 100 Petabytes within the next decade.

The scientific goals and discovery potential of the experiments will only be realized if efficient worldwide access to the data is made possible. Particle physicists are thus engaged in large national and international projects that address this massive data challenge, with special emphasis on distributed data access. There is an acute awareness that the ability to analyze data has not kept up with its increased flow. The traditional approach of extracting data subsets

across the Internet, storing them locally, and processing them with home-brewed tools has reached its limits. Without new modes of data access and of remote collaboration, researchers will not be able to effectively "mine" the intellectual resources represented in their distributed collaborations. Therefore, Caltech researchers are teaming with other scientists from around the globe to explore and implement new ideas in this area that, until now, have only been discussed in a theoretical context.

Building on the results from the Globally Interconnected Object Databases project (ngit2.cithep.caltech.edu/), which was funded by Caltech, CERN, and Hewlett-Packard, CACR is engaged in three new projects: the Particle Physics Data Grid (PPDG), funded by the Department of Energy's Next Generation Internet; Models Of Networked Analysis at Regional Centers (www.cern.ch/MONARC), funded by CERN and Caltech; and Accessing Large Data Archives in Astronomy and Particle Physics (ALDAP), supported by the National Science Foundation Knowledge and Distributed Intelligence program. Development activities include deploying and testing Terabyte-scale databases at a few U.S. universities and laboratories participating in the LHC program, as part of the PPDG (www.cacr.caltech.edu/ppdg) and ALDAP (www.sdss.jhu.edu/~szalay/kdi) projects. In addition to providing a source of simulated events for evaluation of the design and discovery potential of the CMS experiment, the distributed system of object databases will be used to explore and develop effective strategies for distributed data access and analysis at the LHC.

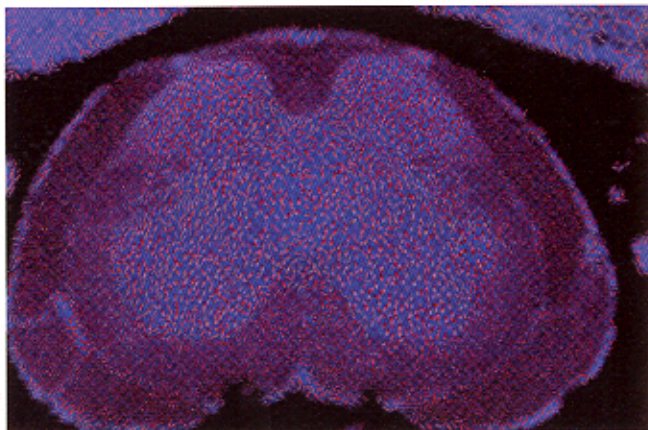
Applying Oil Painting Techniques to Multivalued Scientific Data Visualizations

Advanced computers are employed in a variety of scientific and engineering fields to simulate physical phenomena. The data produced from the simulations needs to be represented, either graphically or otherwise, so that scientists may draw conclusions from the computational results, and make comparisons to experimental and other available data. At Caltech, computational biologists are applying concepts from oil painting to the visual representation of multi-valued biological data. Vector-valued and tensor-valued images are rich sources of information about the physical phenomena they describe, but to be most useful, all of the components of the data must be represented both simultaneously and intuitively. This makes multi-valued scientific data challenging to visualize.

In oil paintings, components of a scene are mapped onto visual characteristics of brush strokes by varying their size, shape, color, texture, opacity, direction, and placement. Artists also convey information by stroke locations relative to one another. In the novel technique developed at Caltech, the same techniques are used to represent the many components of multi-valued data simultaneously, while showing

relationships among them. Painters also build up an image with multiple layers of paint, with each layer representing and encapsulating some components of the data. The lowest layer, or *underpainting*, often roughs out the form of the painting. The layers can be semi-transparent, or sparse, and thus can be built up without obscuring one another.

Caltech researchers combined these techniques, creating an interactive computer graphics system (<http://www.its.caltech.edu/~frasl原因/>) for experimenting with visual brush characteristics



Diffusion tensor image (DTI) visualization using concepts from painting. This stroke image is composited from seven interrelated values, including those layers shown in image #8 on page 5.

and layering to represent components of the data. Initial experiments involved three different data types: diffusion-tensor data showing the pathology of a mouse disease in the spinal cord, 2D vector and tensor measurements of flow over an airfoil, and six-valued magnetic resonance imaging data showing the embryonic development of the mouse brain. The former is represented in the figure (above) and in the three-panel image on page 5.

The images produced by using the Caltech interactive computer graphics system effectively display many data values concurrently. They also qualitatively represent the underlying phenomena intuitively and geometrically, and they emphasize different data values to differing degrees, leading a viewer through the temporal process of understanding the relationships among them. The Caltech research team is continuing to refine the data visualization system and will apply the technique to other data sets in other research areas. The scientists are involved in projects to define the cell division and cell motions that shape the developing embryo and its brain. The ability of this approach to convey time, space, and interrelationships between elements is therefore destined to play a growing role in the design and interpretation of imaging experiments.

Nuclear Matter Calculations on a Lattice

Properties of infinite nuclear matter have been deduced from different approaches. Calculations of finite nuclei offer one opportunity of extrapolating properties of matter such as binding energies and equilibrium densities, but they are strongly influenced by finite size effects. Other, direct calculations of nuclear matter, use sophisticated potentials and aim at the description of ground state properties, but are restricted to approximate methods to address the many-body problem. A third kind of calculation—the so-called lattice gas models—attempts a thermal description of nuclear matter. The goal of these computations is the investigation of a liquid-gas phase transition of nuclear matter

expected to take place at subnuclear densities and low temperatures. They are classical, not quantum mechanical, and in addition, their treatment of parts of the Hamiltonian is inconsistent.

Monte Carlo methods are designed to overcome the dilemmas described above. They enable the many-body problem to be treated exactly, taking into account a full quantum mechanical description. These methods can be used to calculate ground state, as well as thermal properties, of a physical system. While calculations are still computationally expensive, continuing advancements in computer technology make it possible for the first time to attempt a realistic description of nuclear matter.

The nuclear theory group in the Kellogg Radiation Laboratory at Caltech has developed algorithms to calculate thermal and ground state properties of nuclear matter. This involves discretizing space on a lattice, and having nucleons, the constituents of nuclear matter at low temperatures and densities, interact via on-site and nearest-neighbor interactions. The resulting multi-dimensional integrals are calculated with Monte Carlo methods in a stochastic manner. These methods are optimally suited for parallel computers since the computational performance is high: the code runs on each node independently and only a minimal amount of communication between the nodes is required as just one node collects the samples produced by the other nodes.

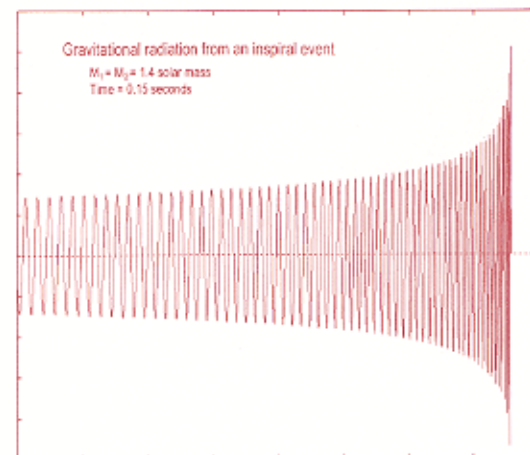
The group used the 256-processor HP Exemplar at CACR. In the first stage of the calculation, the

researchers used a fairly simple nuclear interaction, which will be expanded in future applications. The researchers were able to reproduce energy and basic saturation properties, and produced evidence of a first-order phase transition from an uncorrelated Fermi gas to a clustered system. This was done by observing mechanical and thermodynamical quantities such as compressibility, heat capacity, entropy, and grand potential. Other properties of nuclear matter (such as symmetry energy and sound velocity) are found to be in reasonable agreement with literature. Continuing efforts at Caltech in the further refinement of Monte Carlo methods for these calculations are focused on using bigger lattices and improved interactions.

An Active Digital Library for LIGO

The \$250 million Laser Interferometric Gravitational-Wave Observatory (LIGO) project is designed to detect waves in the fabric of space-time—gravitational waves—using two detectors, one in Washington state and the other in Louisiana. The detectors represent the largest precision optical systems in the world. Each has a pair of 4-km vacuum tubes at right angles, with laser light passing through complex interferometers. The detectors measure distance between the endpoints with incredible precision: if a gravitational wave passes through the system, the tiny changes in geometry can be measured.

In the LIGO project, scientists are searching for gravitational waves generated by three classes of astrophysical processes. An inspiral event is generated when a pair of massive, highly



Signature of an inspiral event on an interferometer

condensed objects, such as neutron stars or black holes, spiral in toward each other and collide. As they orbit, kinetic energy is released through gravitational radiation, causing the orbital period to decrease and the radiation to strengthen. LIGO instruments will detect this as a “chirp” lasting a few seconds, whose signature at the interferometer is shown in the above figure. LIGO researchers are also looking for gravitational waves from supernova events and from the rotation of asymmetric compact objects such as neutron stars, but these are more difficult to find than inspirals. Scientists hope that the raw data from the interferometers contain extremely faint signals deeply buried in noise, requiring supercomputers to search for a variety of inspiral scenarios in the large volume of interferometric data. Once candidate events are found, great care must be taken to eliminate terrestrial effects such as distant earthquakes or spurious electronic signals—in fact, the major reason for having two detectors is to help eliminate local effects.

CACR is collaborating with the LIGO project by providing an “active digital library” to contain the data and processing facilities. Other support provided by CACR includes:

- ♦ *Archiving the data* from the interferometers using the HPSS system. CACR is providing not only data, but also retrieval formats and searching, filtering, and calibration services.
- ♦ Providing *metadata services* that can find which data frames cover a certain time interval, history of the state of these complex instruments, and low-resolution “thumbnail” images of the data frames.
- ♦ Providing *library services* that allow a diverse group of investigators access to the archive, including libraries of possibly related events, such as gamma-ray bursts and seismic events, findings of other investigators, and pointers to related web pages.
- ♦ Providing *computing services* that are connected at high bandwidth to the data archive, to allow the relatively small result of the datamining to be downloaded, rather than the enormous raw dataset.

CACR scientists are also porting the analysis algorithms to high-performance parallel machines and developing visualization software that will allow investigators to see the signatures of candidate and simulated events and their correlations with other data sources.

This visualization depicts the same soliton pseudospherical surface as pictured in image #1 on page 4, but with coloration scheme revealing additional internal structure, as well as a second surface of same type but non-periodic structure. (Bruce Char, Santiago Lombeyda, Ron Perline/Drexel PSE software group)

Computer Science Research at CACR

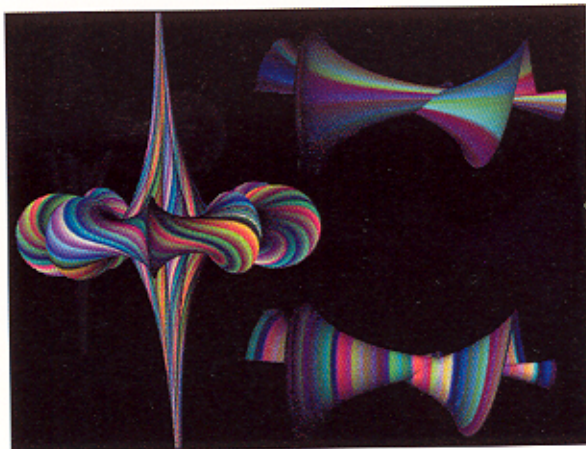
Historically, the central focus of computer science research at CACR has been parallel methods and implementations for scientific simulations, and providing access to advanced parallel computers to conduct realistic experimentation. Those goals remain prerequisites for computational science research, but the modern trends of cheaper and faster hardware and increasingly sophisticated software systems has introduced new challenges. CACR is targeting four major research areas engendered by continuing developments in computational capabilities.

With our current computing technologies, computational power is not always the limiting bottleneck for science and engineering simulations. Increasingly, the hurdles faced by application scientists involve managing and storing large amounts of data, visualizing both the process and results of complex simulations,

controlling remote instruments and gathering data from them, and accessing and effectively using multiple layers of network hierarchy (LAN, WAN, and Grid). CACR scientists are currently creating mechanisms for accessing and managing multi-terabyte-scale data from international collaborations in particle physics and gravity-wave observatories, developing and managing HPSS systems that support the data management problems of high-speed data acquisition, and researching NGI (Next Generation Internet) applications. In collaboration with computer science department faculty, CACR is investigating immersive 3D visualization techniques that use multilevel, hierarchical representations of data.

The ready availability of high-performance PC computers and fast network interconnects has recently made parallel computing on clusters efficient and cost effective. CACR's Beowulf project is developing the systems infrastructure needed to manage and run these clusters, exploring their use for scientific applications, and working with educators to provide Beowulf high-performance computing facilities at the K-12 level.

An important set of challenges in modern high-performance computing research involves how to build and use software on advanced architectures. Problem-solving environments (PSEs) have proven to be an effective mechanism for this. A PSE provides unified access to all the functionality (algebraic, numeric, visualization, data management) needed to solve a large set of problems. Interfaces to the PSE use discipline-specific terminology and ideas,



often through graphical user interfaces. PSEs have typically been crafted individually for different applications, with little reuse of the underlying software framework tools. Caltech's PSEware project focuses on the development of re-usable PSE frameworks that can be extended to a variety of application domains. This work has concentrated on developing a "component architecture" for scientific computing. Component systems provide several advantages for scientific computing: dynamic composition and run-time construction of applications by connecting components that provide stand-alone functionality, distribution of components across a geographically distributed set of machines, components that maintain state and have multiple interfaces, and composition of components in different languages and from different software systems.

Currently, one of the most difficult problems in the computational sciences is determining how to tie together existing research codes from different fields (e.g., fluid dynamics, radiation hydrodynamics, structures, and magnetohydrodynamics) to carry out a simulation with coupled physics. This capability is needed in the Department of Energy ASCI program, in multidisciplinary optimization, in engineering co-design projects, and in recent efforts to merge discrete event simulation with differential equation solvers. These codes are in a state of rapid flux, since they are the main research engines for research groups and are constantly being modified at many levels, ranging from the underlying mathematical model through their interface to the external world. Consequently, a major focus at Caltech and other research

institutions is to develop methods for rapidly connecting solvers using scripting languages such as Python, allowing users in one research area to use the combined software to run an integrated simulation, while concurrently carrying out the organization's primary research programs.

High-Performance Computing in Chemistry: From Very Simple to Very Complex Systems

The numerical solution of the quantum mechanical equations of motion which describe chemical reactions, electron-molecule collisions, and the structure of complex molecules is adding great impetus to the development of chemistry, as is the simulation of the chemical properties of materials. These calculations are very computationally intensive. Caltech chemists have been pioneers in the development of methods appropriate for state-of-the-art high-performance massively parallel computer architectures and have been using such computers intensely and efficiently. In addition to generating new fundamental knowledge, the results of these kinds of calculations permit the modeling of complex chemical systems such as plasmas, novel materials, the atmosphere of the Earth and other planets, and chemical lasers. Such developments in turn enable the generation and improvement of advanced new technologies and materials.

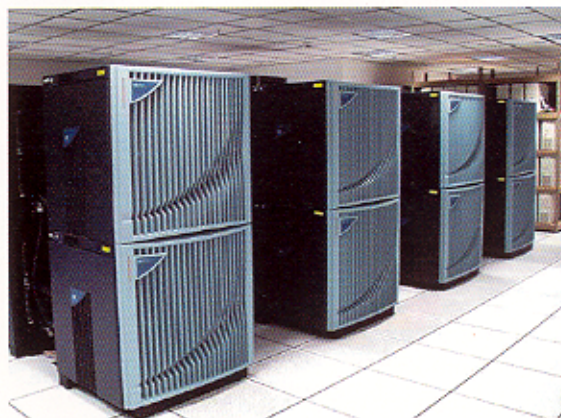
After more than seven decades since the development of modern quantum mechanics, this prophetic statement still poses major theoretical and computational challenges. However, using state-of-the-art high-performance computer architectures and computational methodologies, the quantum mechanical equations describing molecular systems are being solved with sufficient accuracy to provide new insights into these systems, guidance for new kinds of experiments, and much-needed data that often may be experimentally inaccessible.

"...IN THE CONSIDERATION OF ATOMIC AND MOLECULAR STRUCTURE AND ORDINARY CHEMICAL REACTIONS IT IS, INDEED, SUFFICIENTLY ACCURATE IF ONE NEGLECTS THE RELATIVE VARIATION OF MASS WITH VELOCITY AND ASSUMES ONLY COULOMB FORCES BETWEEN VARIOUS ELECTRONS AND ATOMIC NUCLEI. THE UNDERLYING PHYSICAL LAWS NECESSARY FOR THE MATHEMATICAL THEORY OF A LARGE PART OF PHYSICS AND THE WHOLE OF CHEMISTRY ARE THUS COMPLETELY KNOWN, AND THE DIFFICULTY IS ONLY THAT THE EXACT APPLICATION OF THESE LAWS LEADS TO EQUATIONS MUCH TOO COMPLICATED TO BE SOLUBLE. IT THEREFORE BECOMES DESIRABLE THAT APPROXIMATE PRACTICAL METHODS OF APPLYING QUANTUM MECHANICS SHOULD BE DEVELOPED, WHICH CAN LEAD TO AN EXPLANATION OF THE MAIN FEATURES OF COMPLEX ATOMIC SYSTEMS WITHOUT TOO MUCH COMPUTATION."

— P. A. M. Dirac, *Proc. Roy. Soc*
(London) **A 123**:714 (1929).

CACR Computing Facilities

CACR aims to provide new capabilities for computational science and engineering by integrating systems with novel technologies or architectures into facilities that are used by researchers for demanding applications. By combining technologies with new capabilities—such as archives that can store terabytes of data and access them quickly and computers with very large memories—the resulting systems frequently gain unique capabilities and inspire new applications. CACR's computing environment simultaneously provides leading-edge capabilities for CS&E research and experiments with new technologies



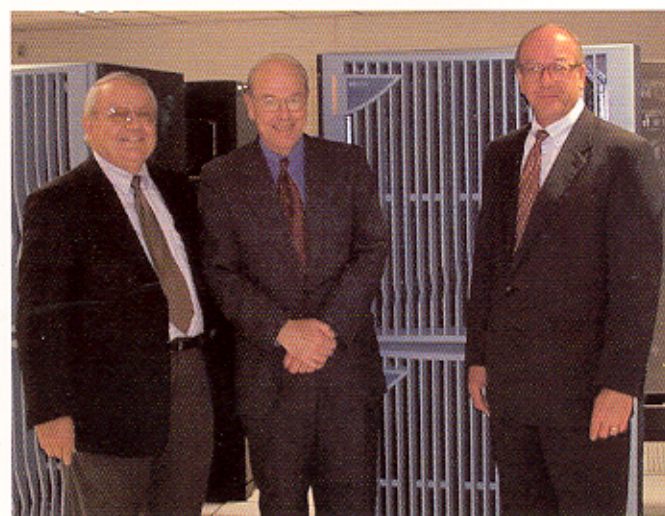
that help define the technical computing environment of the future. Major components of the environment typically include scalable computers with massively parallel architectures; highly parallel disk systems; and high-performance, high-capacity archival

storage system—all connected by local and wide-area networks. The high-speed networks include both production-level and experimental technologies, the latter often operating at record speeds. CACR's computing facilities

are located in the Powell-Booth Laboratory for Computational Science (pictured at upper right) on the Caltech campus.



As scientists study ever more complex phenomena through computer simulation, they often require multidisciplinary approaches, information stored in varied databases, and computing resources that exceed any single computer. Caltech computational projects—which often have data-intensive components, compute-intensive components, or both—rely on very high-speed networks and advanced computing techniques. Consequently, CACR participates in a variety of research and development activities aimed at enhancing our nation's distributed computing capabilities to ensure that Caltech stays at the leading edge in advanced networking capabilities and in metacomputing research. To provide the networking infrastructure to drive Caltech's computational projects, CACR's connections include CalREN-2, the National Transparent Optical Network (NTON), vBNS, ESnet, and High-Speed Connectivity Consortium Network (HSCCnet), among others.



Low Platt, Chairman of the Board of the Hewlett-Packard Company (right)—pictured in the CACR computing facilities, along with James C. T. Pool, CACR Executive Director (left), and Thomas Everhart, president emeritus of Caltech (middle). As part of a longterm partnership with Hewlett-Packard, CACR is installing successive generations of HP's high-end systems (such as the 256-processor Exemplar system pictured above), integrating them into the CACR and NPACI computing environments, and assessing their suitability for large-scale scientific applications, including projects that involve very large scientific databases. The HP systems are shared resources, available for use by Caltech and JPL scientists, as well as researchers nationwide, through NPACI.





**CENTER FOR ADVANCED
COMPUTING RESEARCH**

California Institute of Technology
Mail Code 158-79, Pasadena, CA 91125
(626) 395-6953 • (626) 584-5917 FAX
techpubs@cacr.caltech.edu

For more information on the CACR, see
<http://www.cacr.caltech.edu>

